Self-adapting TV Based Applications

Daniel Costa and Carlos Duarte

LaSIGE and Informatics Department
Faculty of Sciences of the University of Lisbon
Campo Grande, 1749-016 Lisboa, Portugal
thewisher@lasige.di.fc.ul.pt, cad@di.fc.ul.pt

Abstract. The lack of accessibility of ICT applications affects mainly the disabled and/or elderly people, who have a high risk of social exclusion. Using the advantages of adaptive multimodal systems and a well-accepted device like the TV, those difficulties can be surmounted and social exclusion can be stopped. GUIDE intends to simplify the interaction and at the same time not limit the users able to use it, providing multimodal interaction. Also, applications running on GUIDE environment will be able to adapt, on an automated way, to the user's needs and characteristics.

Keywords: Accessibility, Adaptive Multimodal Systems, Dialogue Manager.

1 Introduction

Multimodal interaction has shown a great development in the past years concerning the multiple ways researchers explored to enhance human-computer communication. Being a more "human" way of interacting with computers, using speech, gestures and other "natural" modalities, multimodal interaction is preferred over unimodal interfaces by users [1]. Two main objectives can be set regarding this type of interaction. By one hand, support and accommodate users' perceptual and communicative capabilities; on the other hand integrate computational intelligence in the real world, by offering more natural ways of interaction to humans [2].With this type of interaction, the choices offered to users increase and the complexity of the applications grows. Also, with the increase of computer users we have to deal with different levels of expertise and characteristics [3].

Disabled and/or elderly people are a group at high risk of social exclusion due to the physical, legal, financial, and attitudinal barriers from society that they face in their everyday life. The access to the opportunities offered by society is obviously limited if these cannot be reached by persons with impairments or restricted mobility. A more subtle way of exclusion results from the sensory modalities in which they are presented. Therefore, if the presentation of information has only one modality it will exclude people with impairments in that particular sensory modality. Not being able to use a device or service because its input and output channels support only one modality is a serious restriction of one's everyday life [4]. Adaptive multimodal interfaces, giving the user an optional representation of the same information in more

C. Stephanidis (Ed.): Universal Access in HCI, Part I, HCII 2011, LNCS 6765, pp. 357–364, 2011.

than one sensory mode, and besides adapting automatically to the user's requirements, can compensate to a certain degree for cognitive and sensory impairments. Besides alternative outputs, we have to think about alternative interaction inputs so that no impairment can actually prevent users from using the system.

The European project GUIDE [5] ("Gentle user interfaces for elderly people") is developing a software framework which allows developers to efficiently integrate accessibility features into their applications. GUIDE puts a dedicated focus on the emerging Hybrid TV platforms. These platforms have the potential to become the main media terminals in the users' homes, due to their convenience and wide acceptance. Especially for users of the elderly society, applications such as home automation, audio-visual communication or continuing education can help to simplify their daily life, stay connected in their social network and enhance their understanding of the world. When adapted in the right way, recent advances in human-computer interfaces such as visual gestures, multi-touch as well as speech, or haptics could help to let disabled or elderly users interact with ICT applications in a more intuitive and supportive manner. Nowadays many ICT application implementations simply neglect special needs and lock out a large portion of their potential users. Therefore, this work will help this kind of users to be able to interact with ICT applications in spite of the associated disabilities as it will adapt to the user's needs automatically.

2 Adaptive Multimodal Interfaces

2.1 Adaptation

Adaptation is a "method to increase the usability of an application system in terms of effectivity, efficiency, and ease of use" [6]. We can find many possible cases where interface adaptation would benefit the user interaction: For instance, when the requirements of the user change over time. A user's knowledge and experience evolves over time, from a novice status towards an expert. However the most important reason for having an adaptive system is if the system is used by a heterogeneous user population having different requirements. Users have different needs, preferences, characteristics, knowledge, motivation, goals, etc.

There are many applications that include customization components which allow the user to change the preferences or use templates. Although it is the easiest way of providing adaptation mechanisms from the designer's point of view, it is not so easy for the users. This type of customization done by the user can be very difficult and tedious, and sometimes it isn't even done at all, as pointed out by Berry [7]: "Leaving the customization process to the user, may end up reflecting his beliefs, but not his actual practices, since even domain experts have difficulty describing their true knowledge and biases about a domain". GUIDE has the goal to use self-adaptation, making adaptation the system's responsibility as the users don't have enough knowledge to perform it themselves. However, the user will be capable of adjusting at any time some settings for a better personalization of the system.

2.2 Adaptive Multimodal Systems

The advantages of multimodal interfaces can be better explored by introducing adaptive capabilities. By monitoring the user's interaction and the context of use, the system can adapt automatically, improving its ability to interact with the user, building a user model based on partial experience with that user. These adaptive capabilities are important when dealing with users with different physical and cognitive characteristics, preferences and knowledge. These systems that use multimodal interactions and have the ability to adapt are called Adaptive Multimodal Systems. GUIDE will be used mainly by elderly people, therefore putting together the advantages of multimodality and adaptation to overcome these users' requirements and characteristics is an added value.

Architecture. In general the architecture of an adaptive multimodal interface is composed by the recognizers for the input modalities, the synthesizers for output modalities and between them there is the "Integration committee" [2]. As shown in figure 1, the components for handling the multimodal integration are: a fusion engine (for inputs), a fission module (for outputs), a dialog manager and a context manager.

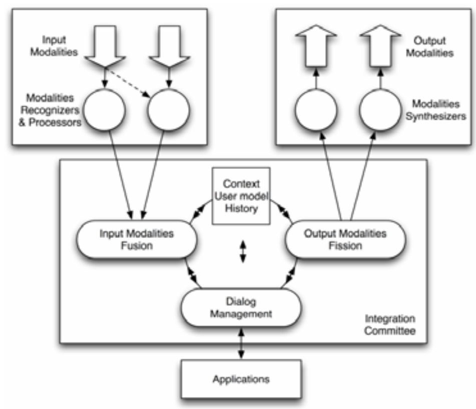

Fig. 1. The architecture of an adaptive multimodal system by Dumas el al [2]

The input modalities, such as speech or gestures, are received by the various recognizers. Then the fusion engine is in charge of giving a common interpretation of those inputs. The interpretation given by the fusion engine is passed to the dialog manager, responsible for identifying the dialog state, the transition to perform, the action to communicate to a given application, and/or the message to return through the fission component. Then, depending on the user profile and context of use, the fission engine returns the message to the user through the most adequate modality or combination of modalities. Finally, the context manager is responsible for communicating any changes on the environment, context and user profile to the other three components, so they can adapt their interpretations.

Context Models. Adaptive systems rely on information stored in models as basis for adaptation. There are different models concerning different types of information to be stored: user model, task model, domain model, dialog model, environment model, etc. The most important model in the adaptive system field is the user model, as the user is the main driver of adaptation in most systems [3]. Being an integral part of any adaptive system, the user model is a representation of knowledge and preferences which the system "believes" that a user (or group of users) possesses [8]. The user model holds the user's characteristics, behavioral patterns, preferences, etc. All this information is important for the adaptation process and for improving the interaction. Therefore, the user model must have continuous maintenance and be always updated.

Knowledge Acquisition Techniques. One of the most relevant aspects of the self-adaptation process is the way the knowledge concerning the users is acquired. According to [3] the process of acquisition of the information needed to build and maintain the user model can be done in an implicit or explicit way. By observing the user's behavior, the interface can obtain information in an implicit way, and use it to enhance the experience and capabilities of the user. This acquisition can be done by observation of direct interaction with the interface, or by the analysis of the information needed by the user. There are two types of information that can be gathered with implicit acquisition: short-term information or long-term information. The short-term information is formed by the current dialog state (e.g., last sentence typed). This information can be used to give immediate assistance in case of a query to the system. The long-term information consists on the interaction history. The explicit acquisition of information for the user model can be done in two ways. The application can ask questions directly to the user regarding the information that is lacking, or the user might be allowed to inspect and possibly make changes to the user model. By asking for information rather than deriving it from other sources, this knowledge acquisition technique can be more precise and reach higher levels of accuracy of the information kept in the user model. However, this involves interruptions on the user task and therefore a more intrusive way of getting the information [3]. An example of the application of these techniques can be seen in [9], where it is described an approach for Personalized Electronic Program Guides to face the information overload and facilitate the selection of the most interesting programs to watch. The system runs on the set-top box and downloads information about the available TV programs from the satellite stream. In order to obtain precise estimates of the individual TV viewer's preferences, their system relies on the management of a hybrid user model that integrates three sources of information: user's explicit preferences; information about viewing preferences of stereotypical TV viewer classes; and user's viewing behavior.

Dialogue Manager Approaches. Bui [10] describes a Dialog Manager (DM) as "the program which coordinates the activity of several subcomponents in a dialogue system and its main goal is to maintain a representation of the current state of the ongoing dialogue". This section will present several approaches to implement a DM.

Finite state models are the simplest models used to develop a dialogue management system. The dialogue structure is represented in the form of state transition networks in which the nodes represent the system's utterances and the transitions between the nodes determine all the possible paths through the network. The Dialogue control is system-driven and all the system's utterances are predetermined. In this approach, both task model and dialogue model are implicit and they are encoded by a dialogue designer. The major advantage of this approach is the simplicity. It is suitable for simple dialogue systems with well-structured task. However, the approach lacks of flexibility, naturalness, and applicability to other domains.

An extension of finite stated-based models, *frame-based model* is developed to overcome the lack of flexibility of the finite state models. The frame-based approach, rather than building a dialogue according to a predetermined sequence of system's utterances, takes the analogy of a form-filling (or slot-filling) task in which a predetermined set of information is to be gathered. This approach allows some degree of mixed-initiative and multiple slot fillings. The task model is represented explicitly and the dialogue model is (implicitly) encoded by a dialogue designer. The frame based approaches have several advantages over the finite state-based approaches: greater flexibility and the dialogue flow is more efficient and natural. However, the system context that contributes to the determination of the system's next action is fairly limited, and more complex transactions cannot be modeled using these approaches.

In [10], Bui refers that *Information state approach* and its extensions are an effort to overcome limitations in finite-state based and frame-based approaches. An information state-based theory of dialogue consists of five main components [11]: A description of informational components (e.g. participants, common ground, linguistic and intentional structure, obligations and commitments, beliefs, intentions, user models, etc.); Formal representations of the above components (e.g., as lists, sets, typed feature structures, records, Discourse Representation Structures, propositions or modal operators within a logic, etc.); A set of dialogue moves that will trigger the update of the information state; A set of update rules, that govern the updating of the information state. And finally, an update strategy for deciding which rule(s) to apply at a given point from a set of applicable ones.

Another extension to the information state approaches is to use *probabilistic techniques* such as (fully observable) Markov Decision Process (MDP) or a Partially Observable Markov Decision Process (POMDP). The idea is to dynamically allow changing of the dialogue strategy and the actions of a dialogue system based on optimizing some kinds of rewards or costs given the current state. This is done by modeling the dialogue as a MDP or as a POMDP. Reinforcement learning is then used to learn an optimal strategy. The actions are the system's responses and questions and the rewards are either defined by the designer (high reward for task completion, low punishment for confirmation and questions and so on) or they are provided by the user who is asked to rate the system at the end of each dialogue.

Plan-based Approaches support a greater complexity to dialogue modeling than the approaches presented in previous sections. Bui [10] explains that these approaches are based on the view that humans communicate to achieve goals, including changes to the mental state of the listener. The dialogue's input is not only considered as a sequence of words but as performing speech acts [12] and it is used to achieve these goals. Usually, the task of the listener is to discover and appropriately respond to the speaker's underlying plan. The plan-based approaches are based on the plan-based theories of communicative action and dialogue which claim that the speaker's speech act is part of a plan and that it is the listener's job to identify and respond appropriately to this plan.

Collaborative Approaches (also known as agent-based dialogue management) are based on viewing dialogues as collaborative process between intelligent agents. Both agents work together to achieve a mutual understanding of the dialogue. Bui [10] points out that unlike the dialogue grammars and plan-based approaches which concentrate on the structure of the task, the collaborative approaches try to capture the motivations behind a dialogue and the mechanisms of dialogue itself.

Machine Learning. Machine learning techniques play an important role in adaptive multimodal interfaces. Many parts of multimodal systems are likely supported by machine learning. For example, the speech, face and gesture recognizers are domains of interest on machine learning. On the self-adaptive systems field, the machine learning can be useful on the user, task and context modeling [2]. Machine learning is the use of algorithms that allow computer programs to automatically improve through experience. Because adaptive user interfaces must learn from observing their user's behavior, another distinguishing characteristic is their need for rapid learning. Dialogue Manager (DM) exists to oversee and control the entire conversational interaction and execute any number of functions. Ultimately, the DM prescribes the next action on each turn of an interaction. Because actions taken by the system directly impact users, the DM is largely responsible for how well the system performs and is perceived to perform by users (i.e. the user experience) [13]. This recommends the use of learning methods that achieve high accuracy from small training sets over those with higher asymptotic accuracy but slower learning rates. On the other hand, the advisory nature of these systems makes this desirable but not essential; an interface that learns slowly will be no less useful than one that does not adapt to the user at all. Still, adaptive interfaces that learn rapidly will be more competitive, in the user's eyes, than ones that learn slowly [14]. An example of a learning method is the reinforcement learning, one the most active research area in artificial intelligence. Reinforcement learning is a computational approach to learning whereby an agent tries to maximize the total amount of reward it receives when interacting with a complex, uncertain environment.

3 GUIDE TV Applications

3.1 Initialization and User Profiling

When the user has contact with GUIDE for the first time it will be presented with an application for initialization and user profiling. The system will ask questions and present tasks so the user preferences and limitations can be perceived and stored and the user model can be built. This is done through a dialogue between the system and the user, which is already predetermined by the application but that should be able to adapt itself based on the knowledge it acquires. These tasks take in account all the elderly related problems such as vision, hearing, motor or cognitive impairments. The application also gathers the user preferences regarding the preferred modalities to interact. All this data is stored in the user model which will be used by the modules that compose the GUIDE adaptation core to understand the user's characteristics so it can apply its techniques for user adaptation.

3.2 Video Conference, Tele-learning, Media Access, Home Automation

GUIDE will provide some applications already installed and prepared to run in its environment. These applications will all run in the GUIDE framework and therefore will be capable of self-adaptation and offer interaction with different modalities. This adaptation is based on the user's characteristics and interaction patterns. For instance, if a person has earing related impairments the system will automatically adapt, presenting the information in other modalities like images and text rather than sound. If a person has cognitive impairments and has difficulties remembering the last screen displayed, then the system will find a way to display the last and current screen at the same time or reduce the steps and screens needed to accomplish a task.

The GUIDE video conferencing application aims at providing inclusive audio-visual communication services for elderly people with ageing-related impairments. Using video conferencing, elderly people shall be enabled to communicate with their families and friends, sharing multimedia content and experiences, being provided with multimodal user interfaces. The goal of the Tele-Learning (TL) Application is to provide to users a means of continuing education by enabling accessible interaction with recorded lecture material. Such scenarios are common in Third-Age Universities, a relatively recent development across Europe that allows older citizens to pursue training in a subject of their interest. Typically, today tele-learning is done on a PC, but in order to make the application accessible to people not necessary experienced with its intricacies, adopting it on a TV environment is considered a major step in making available such a service to a greater audience. The GUIDE Media access application aims at giving easy access to digital TV for elderly and impaired people on a multimodality basis. This application will provide several basic TV services like navigation through other channels while watching a given channel, access to information related to programs on real time, activate subtitles, etc. The home automation application aims at enabling a user to have easy control of his physical home environment through a transparent and distributed user interface that hides the technology in the background and provides the user with a multimodal, intuitive and natural interface to a "reactive home environment". The reactive home environment is equipped with embedded technologies such as sensors, device actuators, input / output channels and software applications that are seamlessly connected in a networked home environment via a user interface framework.

4 Conclusion

In this paper we presented the problems raised by the lack of accessibility on ICT applications and the difficulties that elderly people have to surmount when interacting with such applications. Then GUIDE was described has a system with the goal to overcome those problems, using the TV as the main platform for an adaptive multimodal system. However, to build an adaptive multimodal system that covers all GUIDE's requirements is not an easy task, starting at the architecture and all its components and ending at the adaptation strategies. All of these aspects are complex and have many approaches, therefore which paths to take must be discussed and evaluated. For example, which approach to choose to implement the Dialogue

Manager must take in account the complexity of the applications to be developed or the hardware limitations in terms of processing capabilities. Also protocols must be decided related to the communication between the components and mostly between the adaptation module and the applications.

References

1. Oviatt, S.L.: Advances in Robust Multimodal Interface Design. IEEE Computer Graphics and Applications 23 (September 2003)
2. Dumas, B., Lalanne, D., Oviatt, S.: Multimodal interfaces: A survey of principles, models and frameworks. In: Lalanne, D., Kohlas, J. (eds.) Human Machine Interaction. LNCS, vol. 5440, pp. 3–26. Springer, Heidelberg (2009)
3. Duarte, C.: Design and Evaluation of Adaptive Multimodal Interfaces, PhD Thesis, Faculty of Sciences, University of Lisbon (2007)
4. ETSI. Human Factors (HF): Multimodal interaction, communication and navigation guidelines. In: ETSI Guides and recommendations to promote e-accessibility (2008), http://www.etsi.org/WebSite/Technologies/HumanFactors.aspx
5. Gentle User Interfaces for Elderly People (GUIDE), http://www.guide-project.eu
6. Schneider-Hufschmidt, M., Kuhme, T., Malinowski, U.: Adaptive User Interfaces: Principles and Practice. North-Holland, Amsterdam (1993)
7. Berry, D.C.: The problem of implicit knowledge. Expert Systems 4, 144–151 (1987)
8. Benyon, D., Murray, D.: Applying user modeling to human-computer interaction design. Artificial Intelligence Review 7(3-4), 199–225 (1993)
9. Ardissono, L., Gena, C., Torasso, P., Bellifemine, F., Difino, A., Negro, B.: User Modeling and Recommendation Techniques for Personalized Electronic Program Guides. In: Ardissono, L., Kobsa, A., Maybury, M.T. (eds.) Personalized Information Retrieval (2004)
10. Bui, T.H.: Multimodal Dialogue Management - State of the Art. CTIT Technical Report series No. 06-01, University of Twente (UT), Enschede, The Netherlands (2006)
11. Traum, D., Larsson, S.: The Information State Approach to Dialogue Management (2003)
12. Searle, J.R.: Speech acts: An Essay in the Philosophy of Language. Cambridge University Press, Cambridge (1969)
13. Paek, T., Pieraccini, R.: Automating spoken dialogue management design using machine learning: An industry perspective. Speech Communication 50(8-9), 716–729 (2008)
14. Langley, P.: Machine Learning for Adaptive User Interfaces. KI Kunstliche Intelligenz, 53–62 (1997)